SRA TOOLKIT

- **Prerequisites:** linux or windows subsystem linux https://youtu.be/X-DHaQLrBi8?si=cdt5MvpugeyLSiuA
- Installation process: https://github.com/ncbi/sra-tools https://youtu.be/gs4k3FHkQd0?si=hH7HSw_XrZjIkDER
- Overview of how SRA database works and how to search SRA query

Applications of SRA toolkit :

The Sequence Read Archive (SRA) serves as a vital resource for various applications in genomics, molecular biology, and bioinformatics. Here are some of the primary uses of SRA data:

1. **Genomic Research**: SRA is extensively used in genomic research for studying genetic variations, gene expression, genome assembly, and functional annotation. Researchers can access SRA data to investigate the genetic basis of diseases, evolutionary relationships, population genetics, and comparative genomics.

2. **Transcriptomic Analysis**: SRA data is valuable for transcriptomic studies aimed at understanding gene expression patterns, alternative splicing events, and regulatory mechanisms. Researchers utilize SRA data to analyze gene expression profiles across different tissues, developmental stages, and experimental conditions.

3. **Metagenomics and Microbiome Studies**: SRA provides a wealth of data for studying microbial communities in various environments, including the human gut, soil, oceans, and other ecosystems. Metagenomic analyses using SRA data help in characterizing microbial diversity, identifying novel species, studying community dynamics, and exploring functional pathways.

4. **Epigenomics and Chromatin Biology**: Researchers use SRA data to investigate epigenetic modifications, chromatin accessibility, DNA methylation, and histone modifications. These studies provide insights into gene regulation, cell differentiation, development, and disease mechanisms.

5. **Functional Genomics and Regulatory Networks**: SRA data enables functional genomic studies focused on identifying cis-regulatory elements, transcription factor binding sites, enhancers, and other regulatory sequences. Understanding gene regulatory networks and transcriptional regulation mechanisms is crucial for deciphering complex biological processes.
6. **Disease Research and Biomarker Discovery**: SRA data contributes to research on various diseases, including cancer, infectious diseases, neurodegenerative disorders, and metabolic syndromes. Researchers analyze SRA datasets to identify disease-associated genetic variants, gene expression signatures, diagnostic markers, and therapeutic targets.
7. **Bioinformatics Tool Development and Benchmarking**: Bioinformatics tool developers utilize SRA data for testing and benchmarking algorithms, software tools, and pipelines. SRA datasets serve as standardized benchmarks for evaluating the performance of sequence alignment, variant calling, de novo assembly, and other bioinformatics tools.

Overall, the Sequence Read Archive plays a critical role in advancing scientific knowledge, facilitating data-driven discoveries, and promoting collaboration and data sharing within the research community.